2007

AnimalQTLdb: A Livestock QTL Database Tool Set for Positional QTL Information Mining and Beyond

Zhi-Liang Hu
Iowa State University

Eric Ryan Fritz
Iowa State University

James M. Reecy
Iowa State University, jreecy@iastate.edu

D604–D609 *Nucleic Acids Research, 2007, Vol. 35, Database issue* *Published online 29 November 2006*
doi:10.1093/nar/gkl946

AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond

Zhi-Liang Hu*, Eric Ryan Fritz and James M. Reecy*

Department of Animal Science, Center for Integrated Animal Genomics, Iowa State University, 2255 Kildee Hall, Ames, IA 50011-3150, USA

Received August 15, 2006; Revised October 13, 2006; Accepted October 20, 2006

ABSTRACT

The Animal Quantitative Trait Loci (QTL) database (AnimalQTLdb) is designed to house all publicly available QTL data on livestock animal species from which researchers can easily locate and compare QTL within species. The database tools are also added to link the QTL data to other types of genomic information, such as radiation hybrid (RH) maps, finger printed contig (FPC) physical maps, linkage maps and comparative maps to the human genome, etc. Currently, this database contains data on 1287 pig, 630 cattle and 657 chicken QTL, which are dynamically linked to respective RH, FPC and human comparative maps. We plan to apply the tool to other animal species, and add more structural genome information for alignment, in an attempt to aid comparative structural genome studies (http://www.animalgenome.org/QTLdb/).

INTRODUCTION

Animal genome research has primarily been focused on elucidation of genes that underlie diverse phenotypes using different mapping methods. Rapid progress has been seen in the movement from genome maps, to trait maps and to eventual gene discovery in the past decade. This progress is highlighted by the availability of large numbers of quantitative trait loci (QTL) identified in the mouse, rat, cattle, pigs, chicken, sheep and other species. This large amount of QTL information provides a useful bridge to link genome information with phenotype. However, it must be acknowledged that a bottleneck exists between mapped QTL and gene discovery (1). The ultimate goal of QTL mapping is to identify the genes that underlie polygenic traits and to gain a better understanding of their physiological and biochemical functions. However, success has been limited in utilization of the wealth of QTL information for marker assisted selection (2). This is partly due to the inability to link QTL information to versatile genomic information. One challenge is to find ways to integrate and visualize the QTL data that is produced by different types of QTL analyses, from different laboratories, and with different software tools (3). Development of the PigQTLdb (4) was a significant step towards comparison and consolidation of QTL study results from different laboratories within one species (pig).

Recent advances in radiation hybrid (RH) mapping and genome-sequencing efforts in livestock species has provided useful tools to evaluate and mine QTL information to a higher degree. High resolution, high-density RH maps in pigs (5), chicken (6) and cattle (7,8) contain over 2274, 1228 and 3484 markers, respectively, with an average marker interval of ~1 centimorgan (cM). Furthermore, a majority of the RH markers from these studies served as anchor markers to align with the human genome map, with a comparative coverage of >90% for both cattle and pig maps. The alignment of the RH maps with the human genome map provided excellent tool with which to link livestock QTL regions of interests with all of the genomic information from biomedical studies. The recent availability of cattle, chicken and pig BAC finger printed contig (FPC) maps (9–11) and their emerging genome sequence information provide useful tools for the elucidation of QTL information. Integration of these map resources provides useful tools for positional identification of candidate genes. We have modified our QTLdb to include RH and FPC map information, and added tools to display the human comparative mapping information integrated with QTL locations where applicable.

In this paper, we report the current status of our continued efforts to integrate available structural genomics information for positional QTL information mining in a comprehensive QTL database tool set called *Animal QTLdb*, which has also been revamped to house QTL data from multiple species.

MATERIALS AND METHODS

Data

The QTLdb accepts data curated from published papers, or from private laboratory reports that are in the process for publication. Raw data for >50 parameters/data types is collected. These data includes QTL location (chromosome, location,

*To whom correspondence should be addressed. Tel/Fax: +1 901 759 0643; Email: zhu@iastate.edu
*Correspondence may also be addressed to James M. Reecy. Tel: +1 515 294 9269; Fax: +1 515 294 2401; Email: jreecy@iastate.edu

location span), flanking markers (at suggestive and/or significant test levels), peak markers, test statistics (LOD score, least square means, *P*-values, *F*-values, variance), QTL effects (dominance effect, additive effect), candidate genes, traits and their measurements, etc. The data are organized into related tables in a relational database. The database structure is designed to enable dynamic links to external data resources such as NCBI (Figure 5). The backbone map for displaying marker locations was developed with mapping data mainly from USDA-Meat Animal Research Center (MARC) [for pigs: (12); for cattle: (13)] and Wageningen University [for chicken: (14)]. Markers defining a QTL that were not on the MARC or Wageningen maps were interpolated to form a consensus map with the comparative framework of ArkDB (15).

Data curation

QTLdb curation tools were designed such that a three-tiered data curation structure was implemented, which involves the roles of curators, editors and administrators. This implementation scheme was developed to ensure a smooth workflow and quality control during the entry of new data. The roles of Curators are to gather needed information from literature and scientific reports, translate them into database parameter format of the QTLdb, organize the raw data into an archive for possible 'data debug' and perform the data entry. The roles of Editors are to review the new data entered by Curators, exam the data integrity and scan for errors, including those of grammar, wording and formatting. The roles of Administrators are to manage the database privileges of the Curators and Editors, make new data releases and coordinate data curation and data debug processes.

Software

We used MySQL (version 12.22) as the underlying relational database and Apache 1.3.31 as the worldwide web server. Perl (5.8.5) was employed to program the common gateway interface (CGI) to present and interpret data, synthesize maps, make dynamic links to various data resources and facilitate the interaction between the web portal and the database. The curator/editor tools are also programmed in Perl/CGI. Lincoln Stein's Perl GD library (Lincoln, 2000) was employed to draw QTL map graphs 'on the fly'. All data and tools are hosted on a RedHat Linux server located at Iowa State University.

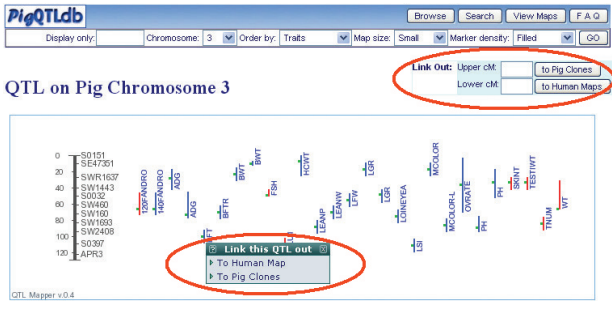

Figure 1. Popup menus and forms (in red circles) were added in the QTLdb 'chromosome view' to link from a QTL region to the RH-human comparative maps or FPC clone maps.

DEVELOPMENTS AND RESULTS

Our progress on development of the QTL database is highlighted with new features for integration of several kind of structural genomic information and inclusion of QTL data from multiple species.

QTL comparisons within a single species

The PigQTLdb has been used to successfully house, compare and manage pig QTL data. A number of enhancements on the database tools were made to efficiently locate, interpret and compare QTL results from different studies. These enhancements include better inter-links between 'browse', 'view maps' and 'search' functions, standardized database summaries, etc. (see Frequently Asked Questions #8 at http://www.animalgenome.org/QTLdb/faq.html).

Expand the QTLdb to include data from more species: potential for comparative QTL analysis

Although there have been a number of similar QTL databases available in various species, our QTLdb has a few unique features over its peers. First, it is designed to house more information to better describe a QTL. Second, it is designed to include trait-associated qualitative variations in the database. Third, it is designed to have the capacity for dynamic cross-platform, cross-internet database links. The value of a good tool is in its practical utility. Toward this end, we have modified the database structure and interface tools, to allow inclusion of QTL data from multiple species. We are now in the process of adding cattle (both dairy and beef) and chicken QTL data. To date, we have collected 630 cattle QTL, 657 chicken QTL and 1287 pig QTL, which represent 89, 112 and 246 cattle, chicken and pig traits, respectively. These QTL were obtained from a total of 180 published journal papers. We continue to actively curate cattle, chicken and pig QTL data into the database and will periodically make database releases to general public.

Trait ontology and its database management

Previously, we introduced a 'trait ontology' to manage diverse pig production traits with a controlled trait vocabulary management system for cross-experiment QTL comparisons [(4); Table 1]. We have now improved our trait ontology management by starting to implement tools to manage directed acyclic graph (DAG) data over the existing simple hierarchical terminology management through an Animal Trait Ontology (ATO) Editor [a.k.a. Collaborative Ontology Builder, COB; (16)]. We realize that the development of Trait Ontology is a long-term effort, which involves active community inputs. We have setup a collaborative platform for contributors in the community to work together on the update and management of the ATO (http://www.animalgenome.org/bioinfo/projects/ATO/).

Genomic mining of QTL region with comparative genome information

High-density RH comparative maps have added new opportunities for researchers to search for genes that underlying a QTL with its fine mapped markers. The MySQL database structure within the QTLdb was modified to house the RH-human

Table 1. A snap shot of the trait ontology organization in a simple tree view of selected pig traits

Trait class	Trai types	Trait names [abbreviation] (custom name)
Meat quality	Fatness	Backfat at first rib [FRIBF] (First rib backfat)
		Backfat at tenth rib [10thRIBF] (Tenth rib backfat)
		Backfat thickness at last rib [LRIBF]
		Backfat at last lumbar [LUMBF] (Lumbar backfat)
	Meat color	Color score [mcolor]
		Color L [mcolor]
		Color A [mcolor]
		Color B [mcolor]
		Hematin pigmentation [mcolor] (Hematin)
	pH	pH 24 h mortem (loin) [pH]
		pH 24 h mortem (ham) [pH]
		pH 45 h mortem [pH]
		pH for semispinalis [pH]
		pH for longissmus dorsi [pH]
	Conductivity	Conductivity 24 h post mortem
		Conductivity 45 min post mortem

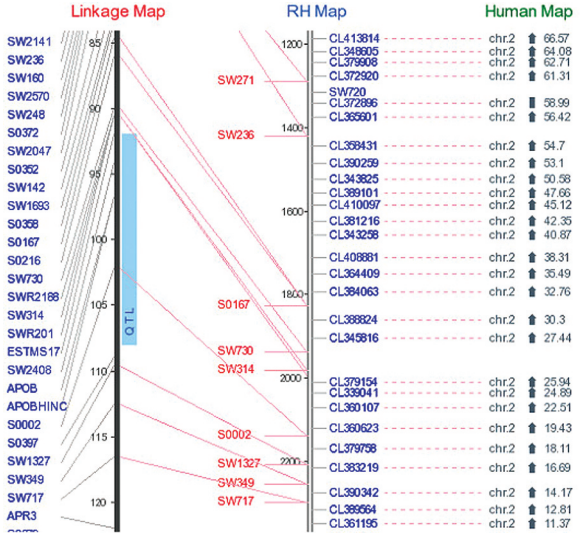

Figure 2. The linked RH-human map showing the region of human chromosome 2, which corresponds to pig chromosome 3 that harbors a QTL for pig basal glucose levels (BGL). Note the interested QTL region is highlighted.

comparative maps for cattle, pigs and chickens. We have successfully incorporated the porcine (5) and cattle (6) RH maps and comparative mapping data into the QTLdb. Chicken RH data (7) is in the process being added. Web tools were programmed to extend the QTL display function to show maps of comparative information alignments (Figure 1). Figure 2 shows an example where a pig QTL on chromosome 2 is aligned to the porcine RH map as well as the human genome using the QTLdb tools. The map visualization is made intuitive for end-users to retrieve such information easily via the QTLdb web viewer. With the aid of this tool, map alignments will be further refined as more sequences become available from the chicken/cattle/pig genome-sequencing project. Toward this end, we also plan to add the chicken/cattle/pig

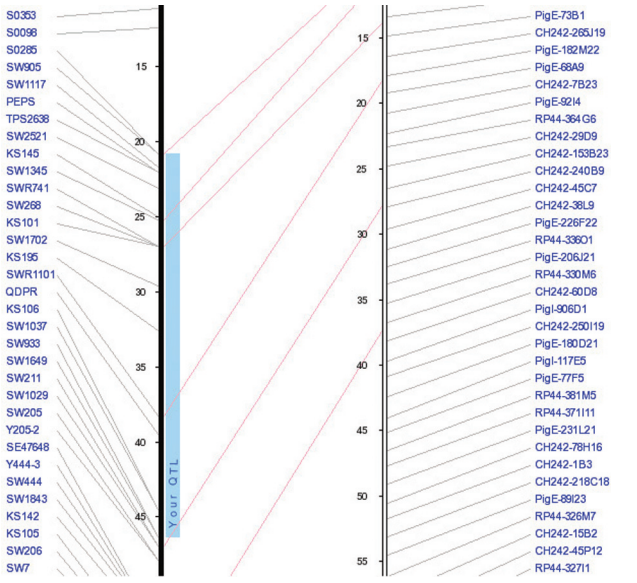

Figure 3. A snap shot of Linkage-FPC map alignment on SSC 8 based on blast information.

genome track information to the QTL maps when they become available.

Aid for selection of positional candidate clones

The dissection of promising genomic regions for candidate genes remains a challenge as the size of such regions normally spans from 2 to 20 cM (on average, 1 cM is ~1 million bases). The recent availability of the 16X-complexity pig BAC fingerprint map provided a new resource for positional cloning (11). We have performed a genome-wide alignment of the pig QTL maps with the Sanger BAC FPC maps by blast analysis of the marker sequences against the BAC end sequences (BES), and subsequently created a map alignment with significant blast hits. To date, we have produced 112 anchor points to link the two maps for all 19 pig chromosomes based on sequence blast matches, and this alignment is continuously improved by adding more anchor sites once new data becomes available, to improve alignment quality. For cattle, we have utilized the BAC–RH map alignment information from the Illinois–Texas high-resolution whole-genome cattle–human comparative map (6).

Tools were created for end-users to search with QTL location information (cM) for underlying BAC clones, and related information for further obtaining the sequences. The map alignments will be further refined as more sequences become available from the respective animal genome-sequencing project. Figure 3 shows a snap shot of porcine SSC 8 alignment with the FPC maps using the QTLdb tools.

Improved curator/editor tools

We have built into the QTLdb a set of data curation tools. The addition of curator/editor tools to the QTLdb made it a more useful tool for the research community by allowing multiple researchers to work together. A number of constraints and checkpoints are enforced during the curation process for data integrity, consistency and error checks. In

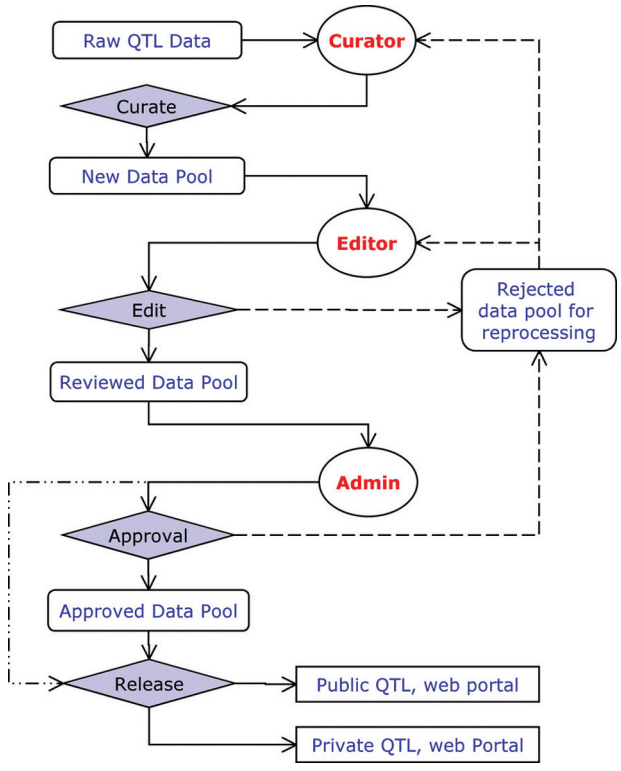

Figure 4. Data flow and curator/editor/administrator work flow built into the QTLdb curator/editor tools.

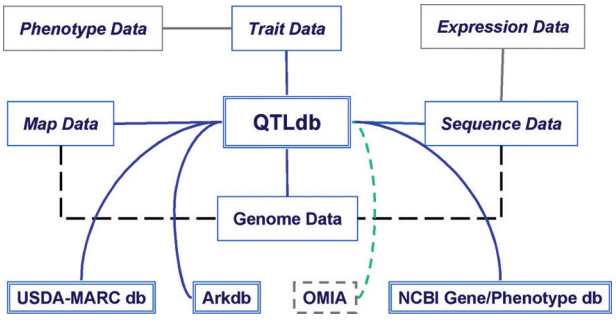

Figure 5. Schematic view of the QTLdb that are designed to enable dynamic links to external data resources (Dotted links: work is on the way to make links; Gray box/lines: data type is planned but does not have actual data yet; Single line boxes: data types; Double line boxes: Databases).

addition, a three-layer quality control mechanism is implemented, with a curator layer, an editor layer and an administrator layer, which helps to minimize the errors, including those of grammar, wording or data formatting. Figure 4 shows the data flow and workflow for curators/editors/administrators that are built into the QTLdb. The tool also helps to keep track of which curator is responsible for which entry, etc. This tool-set uses a server–client model so that it is possible for multiple curators and/or editors from different locations to work together. We have tested the tools with a number of independent curator/editor testers, locally and remotely, and released the tool for public input of new data by registered curators (http://www.animalgenome.org/QTLdb/app.html).

Public data and private data

The QTLdb is also built in such a way that it can house both 'public' and 'private' data. Any new data entered is kept as 'private data' by default until the data owner (curator) wishes to release it to the 'public'. This feature gives general users the flexibility to use the QTLdb as a research tool as well as a data-publishing tool. The 'private data' can be compared to public data through an authenticated user interface. However, the data is not available for comparisons through the public web layers. Being used as a research tool, the QTLdb helps researchers to examine their data against the public data and comparative data in order to draw further inferences with their data set, before they wish to publish the data. In another scenario, if published data is found to be problematic, we can turn that data to 'private' for further scrutinization before release to the public again.

AVAILABILITY

The database contents and tools are all available online at the URL: http://www.animalgenome.org/QTLdb/. The online user-guide can be found at http://www.animalgenome.org/QTLdb/faq.html. For those who wish to enter their own data, this link http://www.animalgenome.org/QTLdb/app.html can be used to apply for access, or login directly once an application is already granted.

DISCUSSIONS AND FUTURE DIRECTIONS

There are several QTL databases in the public domain for rat, mouse and cattle (http://bovineqtl.tamu.edu/) (17–21). The following features make our QTLdb stand out from its peers. First, the Animal QTLdb is designed for and includes data from, multiple species, and is built to have potential for cross-species comparisons through maps or queried by using the trait name/gene name as key words. Secondly, mechanisms are implemented in the Animal QTLdb to allow comparative mapping analysis within species and make it possible to be developed for such comparison between species. Its ability to align pig/cattle/chicken QTL maps to the human genome provides a powerful utility for animal genome researchers to draw inferences from the information-rich human genome. Third, the Animal QTLdb is open to the public for new data entry and existing data management. The ability to share data management privileges will lead to greater participation of scientists from diverse research fields. The general public will have more incentives to contribute their input, and in return, to make the tool more useful. Fourth, by design, the Animal QTLdb allows the most comprehensive QTL description by inclusion of all existing parameters, therefore the QTLdb is more applicable to various situations where diverse analytical methods and parameters may be accommodated. This makes it the most objective comparison tool.

Mapping the location of the QTL is an extremely important first step in QTL identification. While it is clear that location is not sufficient to identify a QTL gene, it is an integral part of the investigation and substantially limits the number of candidate genes that may be identified (22). QTL MatchMaker is

designed to integrate and compare QTL information between the human, the mouse and the rat genomes along with the annotated functional genomics data (19). However, MatchMaker lacks a much-needed tool for graphical display of the map comparisons, and further more, its data input is limited to those already available from other public databases.

We have continued our efforts to utilize existing databases [such as the USDA-MARC database, http://www.marc.usda.gov/genome/; the ArkDB, (15)] and other publicly available data resources (such as PubMed and Gene db of NCBI) by establishing dynamic links that directly point to their query portal for certain data details. We also continued to work closely with other groups such as NCBI, to avoid redundant developmental work. Our efforts on integration of more data types and resources into the Animal QTLdb have helped to outline a genomic data integration pipeline, in which more data types and resources maybe added. We have added RH data, linkage map data and FPC map data to the QTL maps. More data types, such as SNP data, expression data, pathway data and genomic sequence data, etc., may be added as well, when they become available. Our added tools extend the usefulness of the QTLdb and are steps toward an integrated database system for animal genome information.

While we plan to continue our efforts to keep the database up-to-date by adding newly available QTL information, we also plan to include additional animal species for housing their QTL information. In addition to all these, the Animal QTLdb also represents our continued efforts to integrate genome information from distributed databases through Internet. Figure 5 shows a schematic view of the QTLdb that are designed to allow dynamic database links to external data resources. The Animal QTLdb provides a platform for comparative genomics studies across multiple species, with QTL information as a starting point. Our goal is to make a comparative genomics database that any genome characteristics may be an entry point, such as SNPs, candidate genes from another species, interesting phenotypes, annotated expressed sequence tag, etc., to allow users to traverse through genomics information within and across species. This helps users to easily bring related information together for their studies.

ACKNOWLEDGEMENTS

The authors would like to thank Stacey Meyers, Jonathan Beever, Denis Larkin and Harris Lewin from the University of Illinois for providing the pig, cattle RH maps respectively; Sean Humphray, Carol Scott and Jane Rogers from Sanger Institute for their helps in our process to align the pig BAC FPC against the linkage maps; Behnam Abasht and Susan Lamont from Iowa State University for kindly providing the chicken QTL data, Martien Groenen from Wageningen University for their helps with the chicken consensus linkage maps and Max Rothschild as the US Pig Genome Coordinator to bring some needed resources together. In addition, we would like to thank Svetlana Dracheva, Donna Maglott and Wonhee Jang from NCBI for their efforts to incorporate the QTL data through NCBI pipelines. Gratitude also goes to Gary Rohrer of USDA-MARC, and Cathy Ernst of MSU, for their diligent work curating some of the pig QTL data and testing the curator tools. Funding to pay the Open Access publication charges for this article was provided by NRSP-8 National Animal Genome Research Program.

Conflict of interest statement. None declared.

REFERENCES

1. Womack,J.E. (2005) Advances in livestock genomics: Opening the barn door. *Genome Res*., **15**, 1699–1705.
2. Korstanje,R. and Paigen,B. (2002) From QTL to gene: The harvest begins. *Nature Genet*., **31**, 235–236.
3. Sawkins,M.C., Farmer,A.D., Hoisington,D., Sullivan,J., Tolopko,A., Jiang,Z. and Ribaut,J.M. (2004) Comparative Map and Trait Viewer (CMTV): an integrated bioinformatic tool to construct consensus maps and compare QTL and functional genomics data across genomes and experiments. *Plant Mol. Biol*., **56**, 465–480.
4. Hu,Z.-L., Dracheva,S., Jang,W.-H., Maglott,D., Bastiaansen,J., Rothschild,M.F. and Reecy,J.M. (2005) A QTL resource and comparison tool for pigs: PigQTLDB. *Mamm. Genome*, **16**, 792–800.
5. Meyers,S.N., Rogatcheva,M.B., Larkin,D.M., Yerle,M., Milan,D., Hawken,R.J., Schook,L.B. and Beever,J.E. (2005) Piggy-BACing the human genome II. A high-resolution, physically anchored, comparative map of the porcine autosomes. *Genomics*, **86**, 739–52.
6. Everts-van der Wind,A., Larkin,D.M., Green,C.A., Elliott,J.S., Olmstead,C.A., Chiu,R., Schein,J.E., Marra,M.A., Womack,J.E. and Lewin,H.A. (2005) A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. *Proc. Natl Acad. Sci. USA*, **102**, 18526–18531.
7. Morisson,M., Lemiere,A., Bosc,S., Galan,M., Plisson-Petit,F., Pinton,P., Delcros,C., Feve,K., Pitel,F., Fillon,V. *et al*. (2002) ChickRH6: a chicken whole-genome radiation hybrid panel. *Genet. Sel. Evol*., **34**, 521–533.
8. Itoh,T., Watanabe,T., Ihara,N., Mariani,P., Beattie,C.W., Sugimoto,Y. and Takasuga,A. (2005) A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci. *Genomics*, **85**, 413–424.
9. Gustafson,A.L., Tallmadge,R.L., Ramlachan,N., Miller,D., Bird,H., Antczak,D.F., Raudsepp,T., Chowdhary,B.P. and Skow,L.C. (2003) An ordered BAC contig map of the equine major histocompatibility complex. *Cytogenet. Genome Res*., **102**, 189–195.
10. Wallis,J.W., Aerts,J., Groenen,M.A., Crooijmans,R.P., Layman,D., Graves,T.A., Scheer,D.E., Kremitzki,C., Fedele,M.J., Mudd,N.K. *et al*. (2004) A physical map of the chicken genome. *Nature*, **432**, 679–680.
11. Humphray,S., Clark,R., Scott,C., Marron,B., Plumb,R., Bender,C., Camm,N., Daivis,J., Jenks,A., Noon,A., Patel,M., Rogatcheva,M., Milan,D., Chardon,P., Rohrer,G., Nonneman,D., de Jong,P., Meyers,S., Archibald,A., Beever,J., Schook,L. and Rogers,J. (2006) An integrated sequence-ready map of the pig (Sus scrofa domestica) genome. (Dr. Humphray Personal Communication).
12. Rohrer,G.A., Alexander,L.J., Hu,Z., Smith,T.P., Keele,J.W. and Beattie,C.W. (1996) A comprehensive map of the porcine genome. *Genome Res*., **6**, 371–391.
13. Ihara,N., Takasuga,A., Mizoshita,K., Takeda,H., Sugimoto,M., Mizoguchi,Y., Hirano,T., Itoh,T., Watanabe,T., Reed,K.M. *et al*. (2004) Comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Res*., **14**, 1987–1998.
14. Groenen,M.A., Cheng,H.H., Bumstead,N., Benkel,B.F., Elwood Briles,W., Burke,T., Burt,D.W., Crittenden,L.B., Dodgson,J., Hillel,J. *et al*. (2000) A consensus linkage map of the chicken genome. *Genome Res*., **10**, 137–147.
15. Hu,J., Mungall,C., Law,A., Papworth,R., Nelson,J.P., Brown,A., Simpson,I., Leckie,S., Burt,D.W., Hillyard,A.L. *et al*. (2001) The ARKdb: genome databases for farmed and other animals. *Nucleic Acids Res*., **29**, 106–110.
16. Bao,J., Hu,Z.-L., Caragea,D., Reecy,J.M. and Honavar,V. (2006) A ''Tool for Collaborative Construction of Large Biological Ontologies,'' 4th *International Workshop on Biological Data Management (BIDM'06)*, dexa, pp. 191–195, co-located with 17th International Conference on Database and Expert Systems Applications (DEXA'06).
17. Polineni,P., Aragonda,P., Xavier,S.R., Furuta,R. and Adelson,D.L. (2006) The bovine QTL viewer: a web accessible database

of bovine quantitative trait loci. *BMC Bioinformatics*, **7**, 283.
18. Petersen,G., Johnson,P., Andersson,L., Klinga-Levan,K., Gomez-Fabre,P.M. and Stahl,F. (2005) RatMap—rat genome tools and data. *Nucleic Acids Res*., **33**, D492–D494.
19. Kremena,V.S., Song,Q., Zhu,A. and Bottinger,E.P. (2006) QTL MatchMaker: a multi-species quantitative trait loci (QTL) database and query system for annotation of genes and QTL. *Nucleic Acids Res*., **34**.
20. Bogue,M.A. and Grubb,S.C. (2004) The Mouse Phenome Project. *Genetica*, **122**, 71–74.
21. Twigger,S., Lu,J., Shimoyama,M., Chen,D., Pasko,D., Long,H., Ginster,J., Chen,C.F., Nigam,R., Kwitek,A. *et al*. (2002) Rat Genome Database (RGD): mapping disease onto the genome. *Nucleic Acids Res*., **30**, 125–128.
22. Flaherty,L., Herron,B. and Symula,D. (2005) Genomics of the future: Identification of quantitative trait loci in the mouse. *Genome Res*., **15**, 1741–1745.